CLAIMS

We claim:

1. A method for estimating a result size of a Group-By operation comprising:
   (a) calculating a cumulative selectivity based upon aggregation of individual selectivity of each column in a group; and
   (b) multiplying said calculated cumulative selectivity by an input size of said operation.

2. The method of claim 1, wherein the step of calculating a cumulative selectivity includes normalizing a selectivity for each column in said group.

3. The method of claim 2, wherein the step of normalizing a selectivity for each column includes applying a weight factor to said selectivity based upon a relative size of a table in which said column resides.

4. The method of claim 1, wherein the step of calculating a cumulative selectivity is based upon the following mathematical relationship: $S_{ab} = S_a + S_b - (S_a \times S_b)$, wherein S_a is a selectivity of column "a", S_b is the selectivity of column "b", and S_{ab} is a cumulative selectivity of columns "a" and column "b".

5. The method of claim 4, further comprising an iterative application of said mathematical relationship for each additional column in said group.

6. The method of claim 1, wherein the step of calculating a cumulative selectivity includes equivalent columns of said group based upon query predicates.

7. A Group-By operation size estimator comprising:

(a) a selectivity manager adapted to calculate a cumulative selectivity based upon an aggregation of selectivity of an individual column in a group; and

(b) a result size manager adapted to receive said calculated cumulative selectivity from said selectivity manager, and to compute a product of said calculated cumulative selectivity and an input size of said operation.

8. The estimator of claim 7, wherein said selectivity manager is adapted to normalize a selectivity for each column in said group.

9. The estimator of claim 8, wherein normalization of said selectivity includes a weight factor adapted to be applied to said cumulative selectivity calculation.

10. The estimator of claim 9, wherein said weight factor includes a relative size of a table in which said column resides.

11. The estimator of claim 7, wherein said selectivity manager utilizes the following mathematical relationship: $S_{ab} = S_a + S_b - (S_a \times S_b)$, wherein S_a is a selectivity of column "a", S_b is a selectivity of column "b", and S_{ab} is a cumulative selectivity of columns "a" and column "b".

12. The estimator of claim 11, wherein said selectivity manager is adapted to iteratively apply said mathematical relationship for each additional column in said group.

13. The estimator of claim 7, wherein said selectivity manager is adapted to include equivalent columns of said group based upon query predicates.

14. An article comprising:

a computer-readable signal-bearing medium;

means in the medium for calculating a cumulative selectivity of each column in a Group-By operation; and

means in the medium for estimating a result size of said operation based upon said cumulative selectivity.

15. The article of claim 14, wherein the medium is selected from a group consisting of: a recordable data storage medium, and a modulated carrier signal.

16. The article of claim 14, wherein said means for calculating said cumulative selectivity includes means for normalizing a selectivity for each column in said group.

17. The article of claim 16, wherein said means for normalizing said selectivity includes a weight factor based upon of a relative size of a table of said column.

18. The article of claim 14, wherein said means for calculating said cumulative selectivity is inclusive of equivalent columns.

19. A method for estimating a result size of a Group-By operation comprising:

(a) calculating a cumulative selectivity based upon aggregation of individual selectivity of each column in a group, wherein the step of calculating a cumulative selectivity is based upon the following mathematical relationship: $S_{ab} = S_a + S_b - (S_a \times S_b)$, wherein S_a is a selectivity of column "a", S_b is the selectivity of column "b", and S_{ab} is a cumulative selectivity of columns "a" and column "b"; and

(b) multiplying said calculated cumulative selectivity by an input size of said

operation.